

# Perbandingan Metode *Seasonal Autoregressive Integrated Moving Average* (SARIMA) dengan *Support Vector Regression* (SVR) dalam Memprediksi Jumlah Kunjungan Wisatawan Mancanegara ke Bali

Ni Putu Nanik Hendayanti<sup>1</sup>, Maulida Nurhidayati<sup>2</sup>

<sup>1</sup>Sistem Informasi, Institut Teknologi Dan Bisnis STIKOM Bali, Indonesia, <sup>2</sup>Institut Agama Islam Negeri (IAIN) Ponorogo, Indonesia,

<sup>1</sup>nanik@stikom-bali.ac.id, nurhidayatimaulida@gmail.com

DOI: <https://doi.org/10.30812/varian.v3i1>

---

## INFO ARTIKEL

### *Riwayat Artikel:*

Diterima: 11-03-2020

Disetujui: 12-04-2020

---

### *Kata Kunci:*

*Seasonal ARIMA*

*Support Vector*

*Regression (SVR)*

Wisatawan Mancanegara  
Prediksi

---

## ABSTRAK

**Abstrak:** Berbagai sumber pendapatan yang dapat dihasilkan dalam suatu daerah, salah satunya yaitu dalam sektor pariwisata. Seperti halnya sektor yang lain, sektor pariwisata juga memberikan banyak sumbangan bagi pembangunan ekonomi di suatu daerah maupun negara tujuan wisata. Indonesia memiliki banyak tujuan wisata daerah yang sudah terkenal hingga mancanegara salah satunya yaitu Pulau Bali. Bali merupakan daerah yang sudah memiliki kedudukan yang sejajar dengan daerah-daerah tujuan wisata lainnya yang ada di dunia. Sebagai suatu daerah yang sangat berpotensi dalam pengembangan wisata, maka pemerintah memberikan perhatian yang khusus dalam pengembangan pariwisata di Pulau Bali. Maka dari itu, perlu adanya peramalan jumlah kunjungan wisatawan mancanegara ke Bali yang nantinya bisa bermanfaat bagi pemerintah daerah maupun dinas pariwisata. Dalam hal ini, akan digunakan dua metode untuk meramalkan jumlah kunjungan wisatawan mancanegara ke Bali. Adapun metode yang digunakan yaitu *Seasonal ARIMA* dan *Support Vector Regression* (SVR). Hasil peramalan data *out sampel* dengan menggunakan metode SARIMA dan SVR menunjukkan bahwa metode SARIMA memiliki nilai MAPE lebih kecil dari pada SVR. Nilai MAPE metode SARIMA adalah 5,33% sedangkan metode SVR sebesar 19,74%. Begitu juga nilai MSE dan MAE dari metode SARIMA lebih kecil dari metode SVR. Dari Penelitian yang dilakukan dapat disimpulkan bahwa model SARIMA merupakan metode yang lebih baik untuk memprediksi jumlah kunjungan wisatawan mancanegara ke Bali.

**Abstract:** Various sources of income can be produced in an area, one of which is in the tourism sector. As with other sectors, the tourism sector also provides a lot of contributions for economic development in an area or country of tourist destinations. Indonesia has many tourist destinations that are well known to foreign tourists, one of which is the island of Bali. Bali is a region that already has a position parallel to the other tourist destinations in the world. As a very potential area in the development of tourism, the Government gave special attention in the development of tourism in the island of Bali. Therefore, it is necessary to forecasting the number of foreign tourists visit to Bali that can be beneficial to the local government and tourism office. In this case, will be used two methods to predict the number of foreign tourists visit to Bali. The methods used are Seasonal ARIMA and Support Vector Regression (SVR). The results of data-out forecasting samples using the SARIMA and SVR methods indicate that the SARIMA method has a MAPE value smaller than the SVR. The value of Mape methods Sarima is 5.33% while the SVR method is 19.74%. For MSE and MAE values of the SARIMA method is also smaller than the SVR method. From the research conducted it can be concluded that the Sarima model is a better methods to predict the number of foreign tourists visit to Bali.

## A. LATAR BELAKANG

Indonesia merupakan negara kepulauan yang memiliki wilayah terbentang dari Sabang di Aceh hingga Merauke di Papua. Wilayah yang luas ini menjadikan Indonesia memiliki wilayah alam yang mendukung tingkat keberagaman hayati terbesar kedua. Indonesia juga terdiri dari berbagai suku, bahasa, dan agama dengan semboyan “Bhineka Tunggal Ika” yang menjadikan perbedaan bukan sebagai penghalang tetapi sebagai pemersatu bangsa. Keindahan alam dan keanekaragaman budaya yang dimiliki menjadikan Indonesia sebagai salah satu negara di ASEAN yang patut diperimbangkan sebagai tujuan wisata baik pada tingkat regional maupun tingkat internasional (Indrasetyaningih, Damayanti, & Susanto, 2017).

Sektor wisata merupakan sektor yang sangat berpotensi dan menguntungkan untuk dikembangkan sebagai sumber devisa bagi Indonesia. Pariwisata masuk dalam salah satu program pembangunan ekonomi nasional di Indonesia. Dilihat dari penciptaan lapangan kerja, sektor pariwisata mampu menciptakan 1 dari 10 tenaga kerja (Subdirektorat Statistik Pariwisata, 2017). Pertumbuhan kunjungan wisatawan baik domestik maupun mancanegara memberikan dampak positif. Pariwisata juga merupakan salah satu sektor yang menjadi motor penggerak dalam pertumbuhan ekonomi Negara Indonesia.

Salah satu tujuan wisata di Indonesia yang sudah dikenal baik dari dalam maupun luar negeri adalah Provinsi Bali. Provinsi Bali merupakan salah satu wilayah di Indonesia yang memiliki beraneka potensi pariwisata berupa wisata pantai, wisata alam, wisata seni, wisata belanja, dan wisata budaya. Masyarakat Bali dikenal masih memegang teguh tradisi peninggalan nenek moyang menjadikan Bali diminati oleh wisatawan terutama wisatawan mancanegara. Bali memiliki daya tarik yang dapat membuat Indonesia dikenal sebagai salah satu Negara dengan destinasi tempat wisata yang populer. Bagi wisatawan mancanegara, Bali memiliki pantai-pantai dengan ombak yang sangat cocok untuk *surfing* seperti pantai Uluwatu, pantai Kuta serta pantai Dreamland. Hal ini sangat menarik wisatawan yang menyukai tantangan yang dapat memacu adrenalin serta dapat melakukan kegiatan seperti menyelam, *rafting*, maupun *treking*.

Perkembangan pariwisata di Bali sangat ditentukan oleh jumlah kunjungan wisatawan baik wisatawan dalam negeri maupun wisatawan mancanegara. Perkembangan pariwisata akan mempengaruhi sektor publik dan sektor swasta seperti penyediaan infrastruktur yang memadai, transportasi yang dapat menjangkau obyek wisata, promosi destinasi wisata baru, serta layanan yang dapat memudahkan dalam mengakses wisata tersebut. Perkembangan pariwisata yang pesat harus diimbangi pula dengan penyiapan infrastruktur yang memadai agar mampu menangkap semua kondisi yang ada. Pada akhirnya, perkembangan wisata menjadi penggerak roda perekonomian di Bali. Potensi yang besar pada sektor pariwisata menjadikan pariwisata menjadi sektor yang penting untuk dipertimbangkan. Perencanaan jangka panjang perlu dilakukan agar kegiatan yang dilakukan dalam mendukung sektor pariwisata menjadi terarah.

Pentingnya sektor pariwisata terutama bagi devisa Negara yang membutuhkan banyak investasi dalam berbagai aspek, seperti pembangunan infrastruktur, transportasi umum, hotel, serta fasilitas rekreasi di tempat-tempat wisata, yang semua konstruksinya memerlukan waktu yang cukup lama. Perencanaan yang matang perlu dilakukan agar ketika semua infrastruktur telah dikerjakan mampu dimanfaatkan dengan optimal. Selain itu, diperlukan suatu pemikiran yang diharapkan dapat mengantisipasi dampak negatif yang menimbulkan penurunan jumlah wisatawan mancanegara ke Bali. Salah satu cara yang dapat dilakukan untuk mengetahui tingkat kunjungan wisatawan mancanegara pada masa mendatang adalah dengan peramalan. Hal ini dilakukan agar pembangunan yang dilakukan dapat optimal sesuai tingkat kunjungan yang ada. Menurut Herawati (2016) dalam penelitiannya yang berjudul “peramalan kunjungan wisatawan mancanegara menggunakan *generalized regression neural networks*” menyebutkan bahwa peramalan merupakan metode yang dibutuhkan untuk perencanaan kebijakan yang efektif (Herawati, 2016).

Beberapa metode yang untuk melakukan peramalan jumlah kunjungan wisatawan mancanegara ke Bali antara lain metode *Seasonal Autoregressive Integrated Moving Average* (SARIMA) dan *Support Vector Regression* (SVR). SARIMA merupakan salah satu model *Box-Jenkins* yang memanfaatkan perilaku dari data yang diamati dan adanya faktor musiman pada data (Widarjono, 2017). *Support Vector Regression* (SVR)

merupakan metode yang dapat menghasilkan performansi yang bagus karena dapat mengatasi masalah *overfitting* (Santoso, 2007).

Muhammad Hisyam Lee dkk (2012) meneliti tentang Indeks Polusi Udara dengan menggunakan *Seasonal ARIMA* dengan tujuan untuk memodelkan dan meramalkan kualitas udara bulanan di Malaysia pada masa yang akan datang (Lee et al., 2012). Hendayanti dkk (2019) meneliti tentang "Penerapan *Support Vector Regression* (SVR) dalam memprediksi Jumlah Kunjungan Wisatawan Domestik ke Bali". Hasil penelitian tersebut menunjukkan bahwa nilai MAPE yang dihasilkan model SVR baik untuk memprediksi jumlah pengunjung pariwisata (Hendayanti, Suniantara, & Nurhidayati, 2019). Raharyani dkk (2018) meneliti tentang "Implementasi *Algoritme Support Vector Regression* Pada Prediksi Jumlah Pengunjung Pariwisata". Hasil penelitian yang dilakukan menunjukkan bahwa rata-rata nilai MAPE minimum yang diperoleh adalah 9,16% dan nilai MAPE terbaik sebesar 6,98%. Artinya rata-rata selisih antara data aktual dan hasil prediksi sebesar 115 jumlah pengunjung (Raharyani, Putri, & Setiawan, 2018).

Arini dan Nawangsih (2015) meneliti tentang "Peramalan Jumlah Kunjungan Wisatawan Mancanegara (Wisman) ke Bali Tahun 2019: Metode ARIMA". Hasil penelitiannya menunjukkan bahwa hasil ramalan jumlah kedatangan wisman di Bali mencapai 5,07 juta pada tahun 2019 berada dibawah target yang ditentukan oleh Provinsi Bali. Berdasarkan hasil ini, Arini dan Nawangsih menyarankan agar pemerintah memberikan lebih banyak program-program untuk meningkatkan jumlah kunjungan wisatawan agar memenuhi target jumlah kunjungan wisatawan di tahun-tahun yang akan datang (Arini & Nawangsih, 2015). Hadiriyanto dkk (2018) melakukan penelitian tentang "Peramalan Jumlah Wisatawan Mancanegara di Provinsi Bali dengan Menggunakan ARIMA dan Winter". Hasil penelitian menunjukkan bahwa peramalan dengan menggunakan model ARIMA dan winter menghasilkan peramalan yang berbeda secara signifikan dengan selisih yang terpendek yang diperoleh adalah 20.035 dan selisih terpanjang adalah 219.309. Model ARIMA menghasilkan peramalan wisatawan diangka 500.000 sedangkan model winter menghasilkan peramalan jumlah wisatawan dengan pola naik (Hadiriyanto, Darsyah, & Semarang, 2018).

Berdasarkan uraian yang telah dipaparkan sebelumnya, tujuan dari penelitian adalah membandingkan hasil peramalan Jumlah Kunjungan Wisatawan Mancanegara ke Bali dengan metode *Seasonal Autoregressive Integrated Moving Average* (SARIMA) dan *Support Vector Regression* (SVR) untuk mendapatkan hasil terbaik yang selanjutnya dilakukan peramalan jumlah kunjungan wisatawan mancanegara ke Bali pada bulan Januari 2019 hingga Desember 2019.

## B. TINJAUAN PUSTAKA

### 1. *Autoregressive Integrated Moving Average* (ARIMA)

Model *Box-Jenkins* merupakan salah satu teknik peramalan pada model *time series* yang didasarkan pada perilaku variabel yang diamati. Model *Box-Jenkins* secara teknis dikenal sebagai model *Autoregressive Integrated Moving Average* (ARIMA) (Widarjono, 2017). Metode *Box-Jenkins* dalam pemodelan ARIMA terdiri dari beberapa langkah sebagai berikut (Rosadi, 2016)

#### a. *Preprocessing data dan identifikasi model stasioner*

Identifikasi model *time series* digunakan untuk memodelkan sifat-sifat dan perilaku dari data yang diamati. Identifikasi secara sederhana biasanya dilakukan dengan melihat melihat plot dari data. Tujuannya untuk melihat apakah pada data yang diamati terdapat tren, komponen/faktor musiman, non stasioneritas dalam mean dan varian, serta gejala lainnya yang mungkin muncul. Pada proses ini digunakan untuk melihat data yang dapat digunakan pada teknik *preprocessing* untuk membentuk data yang stasioner. Stasioneritas suatu data dibagi menjadi stasioneritas dalam mean dan varian. Stasioneritas dalam mean dapat dilakukan dengan melakukan pengujian *unit root* terhadap data dengan menggunakan uji ADF serta melihat bentuk fungsi autokorelasi (sampel ACF/*Autocorrelation function*) dan PACF dari data. Sedangkan pengujian stasioneritas dalam varian dapat dilakukan dengan menggunakan pengujian *Box-Cox*.

Suatu proses yang stasioner  $\{Z_t\}$  memiliki mean  $E(Z_t) = \mu$  dan varians  $Var(Z_t) = E(Z_t - \mu)^2 = \sigma^2$  yang masing-masing merupakan suatu konstanta kemudian covarian  $Cov(Z_t, Z_s)$  merupakan suatu fungsi dari perbedaan waktu  $|t - s|$ . Covarian antara  $Z_t$  dan  $Z_{t+k}$  didefinisikan sebagai berikut:

$$\gamma_k = Cov(Z_t, Z_{t+k}) = E(Z_t - \mu)(Z_{t+k} - \mu) \tag{1}$$

Korelasi antara  $Z_t$  dan  $Z_{t+k}$  adalah

$$\rho_k = \frac{Cov(Z_t, Z_{t+k})}{\sqrt{Var(Z_t)}\sqrt{Var(Z_{t+k})}} = \frac{\gamma_k}{\gamma_0} \tag{2}$$

dengan  $Var(Z_t) = Var(Z_{t+k}) = \gamma_0$ .

Untuk suatu proses yang stasioner, fungsi autokovarian  $\gamma_k$  dan fungsi autokorelasi  $\rho_k$  memiliki sifat-sifat sebagai berikut:

- 1)  $\gamma_0 = Var(Z_t); \rho_0 = 1$ .
- 2)  $|\gamma_k| \leq \gamma_0; |\rho_k| \leq 1$ .

$\gamma_k = \gamma_{-k}$  dan  $\rho_k = \rho_{-k}$  untuk semua k.  $\gamma_k$  dan  $\rho_k$  adalah fungsi genap yang simetris pada lag k=0. Berdasarkan sifat tersebut, fungsi autokorelasi selalu digambarkan pada lag yang tidak negatif yang selanjutnya disebut sebagai *correlogram* (Wei, 2006).

Autokorelasi parsial digunakan untuk mengukur korelasi antara  $Z_t$  dan  $Z_{t+k}$  setelah menghilangkan atau memisahkan dependensi linier pada variabel  $Z_{t+1}, Z_{t+2}, \dots, Z_{t+k-1}$  terhadap  $Z_{t+k}$  dan dinyatakan sebagai:

$$Corr(Z_t, Z_{t+k} | Z_{t+1}, Z_{t+2}, \dots, Z_{t+k-1}) \tag{3}$$

Autokorelasi parsial dapat diperoleh berdasarkan model regresi dengan variabel dependen adalah  $Z_{t+k}$  dan variabel independen adalah  $Z_{t+k-1}, Z_{t+k-2} + \dots + Z_t$  sehingga model yang terbentuk adalah

$$Z_{t+k} = \phi_{k1} Z_{t+k-1} + \phi_{k2} Z_{t+k-2} + \dots + \phi_{kk} Z_t + e_{t+k} \tag{4}$$

Jika pada saat melakukan tahap *preprocessing* data dan telah dihasilkan data yang stasioner dapat dilanjutkan dengan menentukan bentuk model ARMA (*Autoregressive Moving Average*) dengan jalan menentukan orde p dan q untuk model berdasarkan plot ACF dan PACF.

**b. Estimasi model**

Estimasi terhadap parameter model dilakukan setelah bentuk model dari data ditentukan. Beberapa metode yang digunakan untuk mengestimasi model ARMA serta varian dari residual antara lain metode *Least Square (LS)*, metode *Maksimum Likelihood Estimator (MLE)*, metode *Whittle* maupun metode *Hannan Rissanen*. Dalam pemodelan data *time series*, terkadang dilakukan analisis *overfitting* yaitu suatu analisis yang dilakukan dengan cara mengkaji dan menganalisis model *time series* yang memiliki orde lebih tinggi daripada model yang telah diidentifikasi. Hal ini dilakukan untuk pengujian apakah koefisien yang dihasilkan dari estimasi model signifikan atau tidak. Untuk menguji signifikansi parameter dari model yang diperoleh dilakukan dengan uji t.

**c. Diagnostic Check dan pemilihan model terbaik**

Model yang dihasilkan dari estimasi pada bagian b selanjutnya dilakukan pengujian diagnostik (*diagnostic check*). Pada tahap ini dilakukan konfirmasi kesamaan model sesuai dengan sifat yang dimiliki. Apabila model yang dihasilkan berupa model yang sama, maka data *fitted value* akan dihitung melalui model juga memiliki sifat yang hampir sama dengan data asli. Sehingga residual yang dihitung melalui model mengikuti asumsi dari eror model teoretis (*white noise* dan normalitas dari residual dapat dihilangkan karena tidak sepenting asumsi pertama). Ada tidaknya residual bersifat *White Noise*, dapat dilakukan dengan dua cara, yaitu dilihat dari plot sampel ACF/PACF residual yang terstandardisasi sudah memenuhi sifat-sifat proses *White Noise* dimana nilai mean 0

dan variansi 1 atau melakukan uji korelasi serial. Model yang memenuhi pengujian diagnostik selanjutnya dihitung *mean absolute percentage error (MAPE)*, *mean square error (MSE)*, dan *Median atau Mean Absolut Error (MAE)* untuk digunakan menentukan model terbaik dari beberapa model yang dihasilkan.

#### d. Aplikasi model untuk simulasi, peramalan, dan lainnya

Peramalan sifat-sifat data di masa yang akan datang dilakukan setelah model terbaik diperoleh. Dalam analisis *time series*, data dibagi menjadi dua bagian yaitu data *in sampel* dan data *out sampel*. Data *in sampel* adalah data pemilihan model terbaik dengan langkah-langkah pemodelan sedangkan data *out sampel* adalah bagian dari suatu data yang digunakan untuk memvalidasi keakuratan peramalan dari model terbaik dari data *in sampel*. Model yang baik merupakan model terbaik untuk data *in-sample* dan untuk peramalan, yang dapat diukur dengan data *out sampel*. Adapun jenis dari kebaikan ukuran peramalan seperti ukuran *mean absolute percentage error (MAPE)*, *mean square error (MSE)*, dan *Median atau Mean Absolut Error (MAE)*.

## 2. Model Seasonal ARIMA

Secara umum untuk membuat model seasonal dapat digunakan model berikut (Rosadi, 2012)

### a. Model Seasonal Additive

Pada model *Seasonal Additive* komponen musiman berinteraksi dengan komponen non-musiman dalam model secara *additive*, dinyatakan sebagai SARIMA((p,P),(d,D),(q,Q))<sup>s</sup> bagi proses  $Z_t$  sebagai berikut

$$(1 - a_1 B - \dots - a_p B^p)(1 - B)^d (1 - B_s)^D x_t = (1 + b_1 B + \dots + b_q B^q + \theta_s B^s + \dots + \theta_Q B^{sQ}) \varepsilon_t \quad (5)$$

Dengan

$B$  : operator backward  $(B^j Y)_t = Y_{t-j}$

$p$  : orde non-musiman *autoregressive*

$q$  : orde non-musiman *moving average*

$P$  : orde koefisien *autoregressive* dari komponen musiman multiplikatif

$D$  : order diferensi musiman

$Q$  : order koefisien *moving average* dari komponen musiman multiplikatif

Untuk model pada persamaan 5 komponen musiman digambarkan dengan menggunakan komponen *moving average* orde  $s$ . Komponen musiman dapat juga dinyatakan dengan komponen *autoregressive*, yakni menggunakan model pada persamaan 6.

$$(1 - a_1 B - \dots - a_p B^p - \beta_s B^s - \dots - \beta_p B^{sP})(1 - B)^d (1 - B_s)^D x_t = (1 + b_1 B + \dots + b_q B^q) \varepsilon_t \quad (6)$$

### b. Model Seasonal Multiplikatif

Pada model ini komponen musiman berinteraksi dengan komponen non-musiman dalam model secara multiplikatif (berbentuk perkalian), dapat ditulis sebagai

$$(1 - a_1 B - \dots - a_p B^p)(1 - B)^d (1 - B_s)^D x_t = (1 + b_1 B + \dots + b_q B^q)(1 + \theta_s B^s + \dots + \theta_Q B^{sQ}) \varepsilon_t \quad (7)$$

Di sini komponen musiman digambarkan menggunakan komponen *moving average* orde  $s$ . Komponen musiman dapat juga dinyatakan dengan menggunakan *autoregressive*, yakni

$$(1 - a_1 B - \dots - a_p B^p)(1 - B)^d (1 - B_s)^D x_t = (1 + b_1 B + \dots + b_q B^q)(1 + \theta_s B^s + \dots + \theta_Q B^{sQ}) \varepsilon_t \quad (8)$$

Model multiplikatif ini dapat dinotasikan sebagai model ARIMA (p,d,q)(P,D,Q)<sup>s</sup>.

## 3. Support Vector Regression (SVR)

*Support Vector Regression (SVR)* adalah penerapan dari *Support Vector Machine (SVM)* yang diperkenalkan oleh Cortes dan Vapnik untuk kasus regresi (Cortes & Vapnik, 1995). Tujuan dari *Support Vector Regression (SVR)* yaitu menentukan sebuah fungsi  $f(x)$  sebagai suatu hyperplane (garis

pemisah) berupa fungsi regresi yang mana sesuai dengan semua input data dengan sebuah error. Hasil keluaran SVR berupa bilangan riil dan kontinyu. Rumus training data dapat dilihat pada Rumus 1:

$$\{(x_1, y_1), (x_2, y_2), \dots, (x_\ell, y_\ell)\} \subset \mathcal{X} \times \mathcal{R} \tag{9}$$

input vector dari  $\mathbb{R}^d$  yaitu  $\mathcal{X}$ . Tujuan dari  $\epsilon$ -SV Regression yaitu mencari fungsi  $f(x)$  yang memiliki deviasi maksimal sebesar  $\epsilon$  untuk mendapatkan nilai target  $y_i$  dari semua data *training*, nilai kesalahan akan diterima apabila kurang dari nilai  $\epsilon$  dan ditolak jika nilainya lebih besar dari  $\epsilon$ . Adapun Fungsi linear ditunjukkan pada persamaan dibawah ini:

$$f(x) = \langle w, x \rangle + b \quad w \in \mathcal{X}, b \in \mathbb{R} \tag{10}$$

dimana  $\langle . , . \rangle$  disebut hasil dot product di  $\mathcal{X}$ . Fungsi 1 disebut *flatness* yang bertujuan untuk menemukan nilai kecil dari  $w$ . salah satu cara yang digunakan untuk memperoleh nilai tersebut yaitu meminimalkan bentuk *Euclidean*  $\|w\|^2$  adapun permasalahan *convex optimization* dapat di lihat pada persamaan dibawah ini:

$$\begin{aligned} \text{Minimize} \quad & \frac{1}{2} \|w\|^2 \\ \text{Subject to:} \quad & y_i - \langle w, x \rangle - b \leq \epsilon \\ & \langle w, x \rangle + b - y_i \leq \epsilon \end{aligned} \tag{11}$$

Konstanta  $C > 0$  digunakan untuk memastikan *trade of* dengan *flatness* dari fungsi  $f$  dan besarnya tingkat deviasi kesalahan dari batas  $\epsilon$  yang dapat ditoleransi. Jika dilihat berdasarkan ilmu matematika permasalahan pada SVR yaitu *Convex Linear Programing NLP optimization problem* yang berarti masalah untuk meminimalkan fungsi kuadratik untuk diubah menjadi sebuah batasan. Batasan- ini bisa diselesaikan dengan *dual problem formulation* dimana tahap menyelesaikan kondisi dual problem bisa menggunakan persamaan *Lagrange Multiplier*. Adapun rumus yang digunakan dapat dilihat sebagai berikut:

$$f(x) = \sum_{i=1}^{\ell} (\alpha_i - \alpha_i^*) \cdot \langle x_i, x \rangle + b \tag{12}$$

Rumus pada persamaan 12 merupakan fungsi menghitung permasalahan linear. Sedangkan untuk kasus *non linear* nilai  $x_i$  dan  $x$  pertama-tama ditransformasikan kedalam *space feature* yang berdimensi tinggi dengan cara memetakan vektor  $x_i$  dan  $x$  ke dalam fungsi  $\Phi$ . Fungsi  $\Phi$  merupakan fungsi kernel. Persamaan akhirnya dapat dilihat pada persamaan 13 sebagai berikut:

$$f(x) = \sum_{i=1}^{\ell} (\alpha_i - \alpha_i^*) \cdot K(x_i, x) + b \tag{13}$$

#### 4. Fungsi Kernel

Menurut Budi Santoso, data mining atau *machine learning* merupakan teknik yang dikembangkan dengan berdasarkan asumsi kelinieran yang mengakibatkan algoritma yang dihasilkan terbatas untuk kasus-kasus yang linier. Dengan metode kernel suatu data  $x$  di *input space* dipetakan ke *feature space* dengan dimensi yang lebih tinggi melalui  $\varphi$  (Santoso, 2007).

$$\varphi : x \rightarrow \varphi(x) \tag{14}$$

#### 5. Ukuran Error

Menurut Makridakis dkk (1999) untuk menilai performansi suatu peramalan dapat dilakukan dengan menggunakan ukuran-ukuran error sebagai berikut (Makridakis, Wheelwright, & McGee, 1999):

a. Mean Absolut Percentage Error (MAPE):

$$MAPE = \left( \frac{1}{n} \sum_{i=1}^n \left| \frac{e_i}{y_i} \right| \right) \times 100\% \tag{15}$$

b. Mean Squared Error (MSE):

$$MSE = \frac{1}{n} \sum_{i=1}^n e_i^2 \tag{16}$$

c. Mean Absolut Error (MAE):

$$MAE = \frac{1}{n} \sum_{i=1}^n |e_i| \quad (17)$$

Dimana  $e_i$  adalah selisih  $y_i$  dan  $\hat{y}_i$ ,  $y_i$  adalah data aktual dengan  $i=1, 2, \dots, n$  dan  $\hat{y}_i$  adalah data peramalan dengan  $i=1, 2, \dots, n$ .

### C. METODE PENELITIAN

Metode yang digunakan dalam penelitian ini adalah *Seasonal Autoregressive Integrateg Moving Average* (SARIMA) dan *Support Vector Regression* (SVR). Tujuannya adalah untuk memprediksi jumlah kunjungan wisatawan ke Bali. Data yang digunakan merupakan data bulanan dari BPS Bali periode Januari 2004 sampai dengan Desember 2019. Data selanjutnya dibagi menjadi data *in sampel* dan data *out sampel*. Data *in sampel* dimulai dari Januari 2004 sampai Desember 2018 sedangkan data *out sampel* adalah data dari Januari 2019 hingga Desember 2019. Langkah-langkah analisis dengan SARIMA dan SVR sebagai berikut:

1. Deskripsi Data
2. Analisis data dengan SARIMA

Analisis pada dengan SARIMA meliputi *Preprocessing* data dan identifikasi model stasioner, estimasi model, *Diagnostic Check* dilanjutkan pemilihan model terbaik. Selanjutnya dilakukan peramalan dengan model yang diperoleh.

3. Analisis data dengan SVR

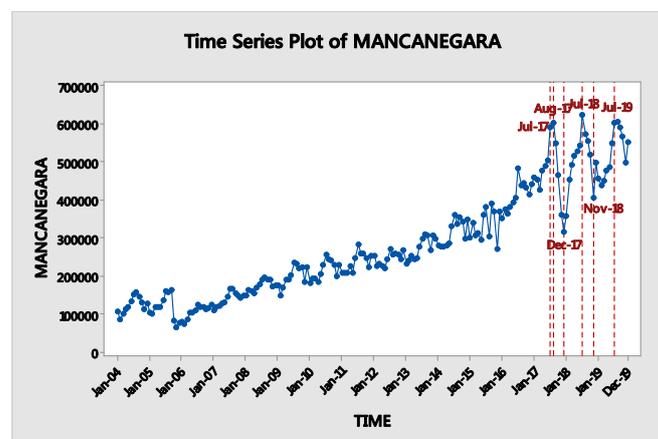
Analisis data dengan SVR dengan tahapan transformasi, membagi data menjadi data *training* dan data *testing*, menentukan tipe kernel, parameter, dan pemilihan model terbaik melalui *grid search Optimiation (GMO)* data *training*, dilanjutkan melakukan peramalan terhadap data *testing*.

4. Perbandingan Metode SARIMA dan SVR
5. Peramalan periode selanjutnya

### D. HASIL DAN PEMBAHASAN

#### 1. Deskripsi data

Penelitian ini menggunakan data jumlah kunjungan wisatawan mancanegara ke Bali dari Januari 2004 sampai dengan Desember 2019. Jumlah kunjungan wisatawan mancanegara ini diperoleh dari Badan Pusat Statistik Bali melalui laman [www.bali.bps.go.id](http://www.bali.bps.go.id). Adapun data tersebut dapat dilihat pada Gambar 1.



Gambar 1. Plot Jumlah Kunjungan Wisatawan Mancanegara Ke Bali

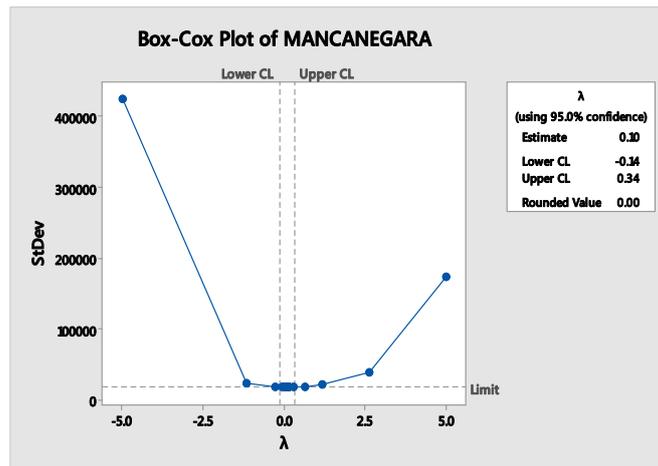
Gambar 1 merupakan perkembangan jumlah kunjungan wisata mancanegara ke Bali dari Januari 2004 hingga Desember 2019. Dilihat dari Gambar 1, jumlah kunjungan wisatawan mancanegara memiliki tren musiman yaitu cenderung naik pada bulan Juli dan Agustus dan tidak jarang mengalami

penurunan di bulan November dan Desember. Berdasarkan data kunjungan wisatawan 3 tahun terakhir diketahui bahwa bulan Juli 2018 menunjukkan jumlah kunjungan paling besar dibandingkan bulan yang lain yaitu sebesar 624366 jiwa dan paling rendah pada bulan Desem 2017 yaitu sebesar 315909 jiwa.

**2. Analisis data dengan SARIMA**

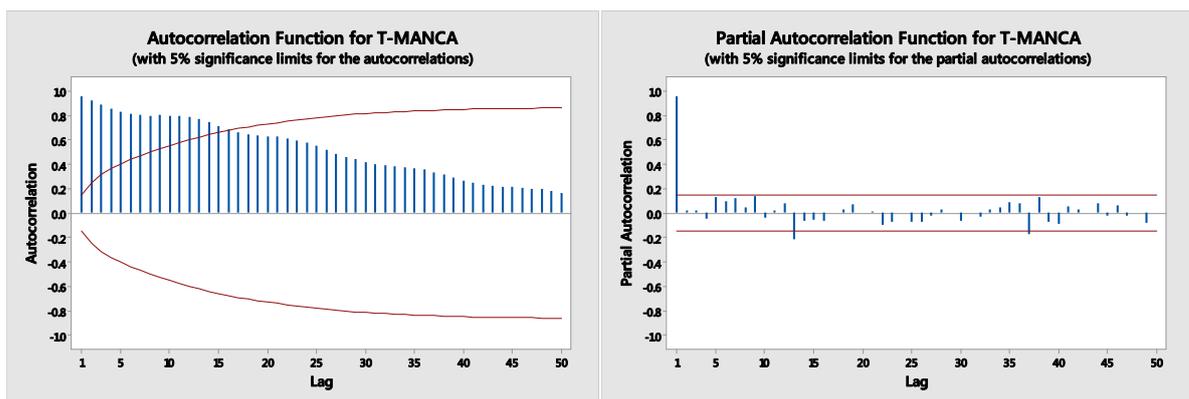
**a. Preprocessing data dan identifikasi model stasioner**

Pada tahap ini dilakukan pengujian stasioneritas baik dalam mean maupun dalam varian. Pengujian stasioneritas dalam varian dengan pengujian *Box-Cox*. Hasil pengujian *Box-Cox* ditunjukkan pada Gambar 2.



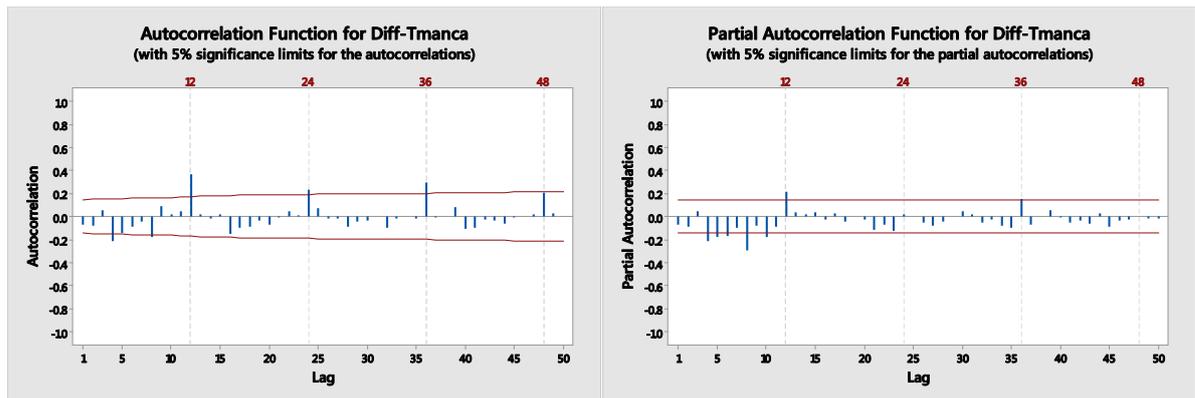
Gambar 2. Hasil Pengujian *Box-Cox*

Hasil pengujian *Box-Cox* pada Gambar 2 menunjukkan bahwa nilai estimasi  $\lambda$  yang diperoleh adalah 0,00 yang berarti bahwa transformasi harus dilakukan dengan fungsi Logaritma. Data hasil transformasi selanjutnya diuji stasioneritasnya dalam mean dengan menggunakan *ACF* dan *PACF* seperti ditunjukkan pada Gambar 3.



Gambar 3. Plot *ACF* dan *PACF* data Transformasi

Hasil plot *ACF* dan *PACF* pada Gambar 3 menunjukkan bahwa *ACF* turun lambat dan *PACF cuts off* pada lag 1 sehingga data transformasi tidak stasioner dalam mean sehingga perlu dilakukan *differencing* untuk menstasionerkan data. Hasil plot *ACF* dan *PACF* data *differencing* ditunjukkan pada Gambar 4.



Gambar 4. Plot ACF dan PACF Data Differencing

Gambar 4 menunjukkan bahwa plot ACF mengalami *cuts off* pada lag 12, 24, dan 36 sedangkan plot PACF *cuts off* pada lag 12 sehingga data jumlah kunjungan wisatawan mancanegara ke Bali mengikuti model SARIMA(0,1,0)(1,0,1)<sup>12</sup>. Pada penelitian ini dicobakan juga model dengan orde yang lebih tinggi yaitu model SARIMA(1,1,0)(1,0,1)<sup>12</sup>.

**b. Estimasi model dan pemilihan model terbaik**

Hasil estimasi parameter model SARIMA ditunjukkan pada Tabel 1.

Tabel 1. Hasil Estimasi Model SARIMA

Model	Type	Coef	P-value
SARIMA(0,1,0)(1,0,1) <sup>12</sup>	SAR 12	0,9944	0,000
	SMA 12	0,8933	0,000
	Const	-0,00017	0,872
SARIMA(1,1,0)(1,0,1) <sup>12</sup>	AR 1	-0,1582	0,037
	SAR 12	0,9950	0,000
	SMA 12	0,8700	0,000
	Const	0,000046	0,966

Tabel 1 menunjukkan bahwa model SARIMA(0,1,0)(1,0,1)<sup>12</sup> dan model SARIMA(1,1,0)(1,0,1)<sup>12</sup> memiliki parameter yang signifikan karena nilai P-value yang dihasilkan kurang dari 0,05 untuk semua parameter dalam model. Karena semua parameter model SARIMA(0,1,0)(1,0,1)<sup>12</sup> dan model SARIMA(1,1,0)(1,0,1)<sup>12</sup>, maka tidak ada parameter yang dikeluarkan dan dapat dilanjutkan dengan *diagnostic check* untuk memastikan model yang diperoleh dari hasil estimasi memenuhi asumsi *white noise* dengan menggunakan pengujian *Ljung-Box*.

Tabel 2. Hasil Pengujian White Noise

Model	Lag	Chi-square	P-value
SARIMA(0,1,0)(1,0,1) <sup>12</sup>	12	16,7	0,054
	24	27,2	0,165
	36	36,0	0,332
	48	44,7	0,485
SARIMA(1,1,0)(1,0,1) <sup>12</sup>	12	15,2	0,056
	24	23,8	0,252
	36	31,9	0,472
	48	39,0	0,687

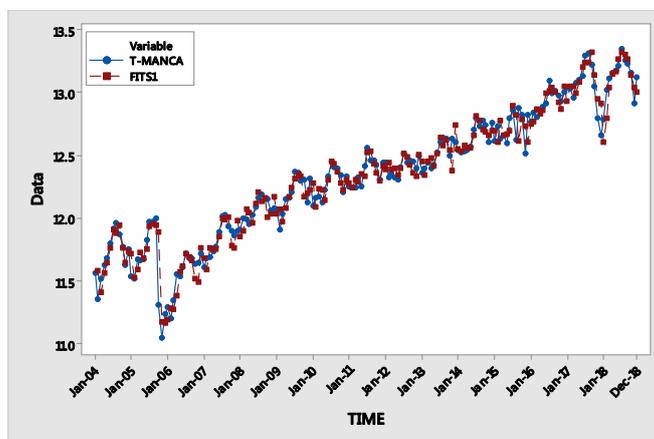
Berdasarkan Tabel 2 diketahui bahwa semua statistik Q untuk lag 12, 24, 36, dan 48 untuk kedua model SARIMA tidak signifikan sehingga kedua model SARIMA bersifat *white noise*. Untuk menentukan model terbaik dari kedua model SARIMA yang dimiliki dilakukan dengan melihat nilai

performa masing-masing model dengan menggunakan nilai MAPE, MSE serta MAE. Semakin kecil nilai MAPE, MSE serta MAE menunjukkan bahwa model tersebut merupakan model yang paling baik untuk digunakan. Hasil uji performa model SARIMA ditunjukkan pada Tabel 3.

**Tabel 3.** Hasil Uji Performa Model SARIMA

Uji Performa Model	SARIMA(0,1,0)(1,0,1) <sup>12</sup>	SARIMA(1,1,0)(1,0,1) <sup>12</sup>
MAPE	6,90%	6,85%
MSE	623x10 <sup>6</sup>	619x10 <sup>6</sup>
MAE	17144,25	16982,58

Berdasarkan Tabel 3 diketahui bahwa nilai MAPE, MSE, dan MAE model SARIMA(1,1,0)(1,0,1)<sup>12</sup> lebih kecil dibandingkan dengan model SARIMA(0,1,0)(1,0,1)<sup>12</sup> sehingga model SARIMA yang terbaik dan digunakan untuk analisis lebih lanjut adalah model SARIMA(1,1,0)(1,0,1)<sup>12</sup>. Gambar 5 merupakan plot data transformasi dan data prediksi model SARIMA(1,1,0)(1,0,1)<sup>12</sup>.



**Gambar 5.** Plot Data kunjungan wisata dan Model SARIMA(1,1,0)(1,0,1)<sup>12</sup>

Hasil prediksi yang dilakukan dengan model SARIMA pada Gambar 5 menunjukkan bahwa model SARIMA mampu mengikuti pola dari data yang dimiliki. Dalam hal ini adalah data transformasi jumlah kunjungan wisatawan mancanegara ke Bali.

**c. Peramalan data out sampel model SARIMA**

Setelah diperoleh model SARIMA yang optimal, selanjutnya dilakukan peramalan 12 bulan ke depan dan diperoleh nilai MAPE=5,33%, MSE=1,056x10<sup>9</sup> dan nilai MAE sebesar 26441,57. Nilai performa ini lebih tinggi dibandingkan dengan nilai performa data in sampel.

**3. Analisis data dengan SVR**

**a. Transformasi data**

Transformasi data dilakukan untuk memastikan bahwa varian dari data tidak terlalu besar. Hasil transformasi data ditunjukkan pada Gambar 2. Hasil tersebut menunjukkan bahwa data memiliki varian yang cukup besar sehingga dilakukan transformasi Logaritma untuk memperkecil varian yang dimiliki.

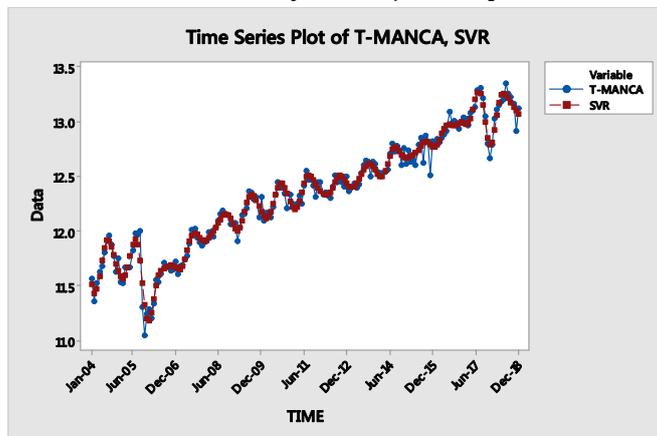
**b. Membagi data menjadi data training dan data testing**

Data hasil transformasi selanjutnya dibagi menjadi data training dan data testing. Data training menggunakan data bulan Januari 2004 hingga Desember 2018 sedangkan data testing menggunakan data bulan Januari 2019 hingga Desember 2019.

**c. Menentukan tipe kernel, parameter, dan pemilihan model terbaik**

Estimasi model SVR dalam penelitian ini menggunakan software R dengan package e1071. Pemodelan data dilakukan dengan menggunakan data training. Pada penelitian ini, kernel yang

digunakan adalah RBF (*radial basis function*) dengan nilai parameternya yaitu *epsilon* adalah 0,1; *cost*=127 serta nilai *gamma*=63,5 merupakan nilai yang diperoleh dengan metode *Grid Search Optimization (GSO)* dengan nilai *performance* sebesar 0,01131. Model yang diperoleh selanjutnya digunakan untuk prediksi data dan hasilnya ditunjukkan pada Gambar 6.



Gambar 6. Prediksi Plot dengan SVR

Gambar 6 merupakan hasil prediksi data *training* berdasarkan model terbaik yang diperoleh dari metode *Grid Search Optimization (GSO)*. Hasil tersebut menunjukkan bahwa prediksi yang diperoleh mengikuti pola dari data *training*. Hasil perormna model menggunakan perhitungan MAPE, MSE, dan MAE ditunjukkan pada Tabel 4.

Tabel 4. Hasil Uji Performa Model Data *Training*

Uji Performa Model	Nilai
MAPE	5,29%
MSE	354x10 <sup>6</sup>
MAE	13331,44

Tabel 4 menunjukkan bahwa MAPE yang diperoleh pada model *training* adalah 5,29% dengan MSE adalah 345x10<sup>6</sup> dan MAE sebesar 13331,44.

**d. Melakukan peramalan terhadap data *testing***

Setelah diperoleh model SVR yang optimal pada data *training* selanjutnya nilai parameter tersebut diimplementasikan pada data *testing*. Hasil peramalan menunjukkan bahwa nilai performa data *testing* adalah MAPE=19,74%, MSE=1,647x10<sup>10</sup> dan nilai MAE sebesar 105903,13. Nilai performa ini lebih tinggi dibandingkan dengan nilai performa data *training*.

**4. Perbandingan metode SARIMA dan SVR**

Untuk membandingkan metode yang memberikan hasil yang terbaik dilakukan menggunakan uji performansi model yaitu MAPE, MSE, dan MAE pada data *out sampel/testing*. Berdasarkan ketiga uji ini selanjutnya akan diketahui model yang paling baik. Hasil uji performansi model ditunjukkan pada Tabel 5.

Tabel 5. Hasil Uji Performa Model SARIMA dan SVR data *testing*

Uji Performa Model	SARIMA(1,1,0)(1,0,1) <sup>12</sup>	SVR
MAPE	5,33%	19,74%
MSE	1,056x10 <sup>9</sup>	1,647x10 <sup>10</sup>
MAE	26441,57	105903,13

Berdasarkan Tabel 5 diketahui bahwa nilai MAPE untuk model SARIMA lebih rendah dibandingkan motode SVR. Begitu juga dengan MSE dan MAE motode SARIMA lebih rendah

dibandingkan metode SVR sehingga dapat disimpulkan bahwa model SARIMA merupakan model terbaik untuk memodelkan data kunjungan wisatawan mancanegara ke Bali.

### 5. Peramalan periode selanjutnya

Pada tahap ini merupakan hasil peramalan jumlah kunjungan wisatawan mancanegara ke Bali untuk bulan Januari sampai dengan Desember tahun 2019 dengan metode SARIMA(1,1,0)(1,0,1)<sup>12</sup>. Hasil peramalan ditunjukkan pada tabel 6.

**Tabel 6.** Prediksi Jumlah Kunjungan Wisatawan Mancanegara Ke Bali dengan SARIMA(1,1,0)(1,0,1)<sup>12</sup>

Bulan	Jumlah Kunjungan
Januari 19	472637
Februari 19	490938
Maret 19	500063
April 19	521288
Mei 19	528113
Juni 19	575354
Juli 19	647230
Agustus 19	614339
September 19	616373
Oktober 19	567872
November 19	496729
Desember 19	545815

Tabel 6 menunjukkan prediksi jumlah kunjungan wisatawan mancanegara ke Bali tahun 2019 mengalami peningkatan paling besar pada bulan Juli 2019 dengan jumlah kunjungan sebesar 647.230 dan turun kembali pada bulan-bulan berikutnya serta kunjungan paling kecil terjadi pada bulan Januari 2019 sebesar 472.637 kunjungan.

### E. SIMPULAN DAN SARAN

Kesimpulan yang dapat diambil dari hasil penelitian yang telah dilakukan adalah sebagai berikut:

1. Penentuan model terbaik dilakukan dengan menggunakan performa dari MAE, MSE, dan MAPE. Semakin kecil nilai MAE, MSE, dan MAPE maka semakin baik hasil estimasi yang diperoleh.
2. Berdasarkan hasil yang diperoleh dengan menggunakan data *out sampel* atau data *testing* diketahui bahwa nilai MAPE, MSE, dan MAE yang dihasilkan pada model SARIMA lebih kecil dibandingkan dengan model SVR sehingga model SARIMA lebih sesuai digunakan untuk prediksi jumlah kunjungan wisatawan mancanegara ke Bali.
3. Peramalan jumlah kunjungan *wisatawan* mancanegara ke Bali untuk bulan Januari sampai dengan Desember tahun 2019 menunjukkan bahwa jumlah wisatawan mancanegara ke Bali mengalami peningkatan paling besar pada bulan Juli 2019 dengan jumlah kunjungan sebesar 647.230 dan turun kembali pada bulan-bulan berikutnya serta kunjungan paling kecil terjadi pada bulan Januari 2019 sebesar 472.637 kunjungan.

Adapun saran yang penulis dapat sampaikan dari penelitian ini adalah pada penelitian selanjutnya dapat dicoba membandingkan kembali metode SARIMA dengan model SVR dengan menentukan terlebih dahulu parameter (*epsilon*, *gamma*, dan *cost*) yang sesuai untuk metode SVR. Semakin sesuai parameter yang diambil akan semakin mendekati kondisi sebenarnya dalam data.

### UCAPAN TERIMA KASIH

Saya ucapkan terima kasih untuk ITB STIKOM Bali yang sudah memberikan kesempatan kepada saya dalam mengajukan penelitian ini dan sudah mendanai penelitian ini serta keluarga dan teman-teman yang sudah memberikan doa dan motivasi dalam melancarkan penulisan ini.

## REFERENSI

- Arini, P. S., & Nawangsih, E. (2015). Peramalan Jumlah Kunjungan Wisatawan Mancanegara (Wisman) ke Bali Tahun 2019: Metode ARIMA. *Jurnal Ekonomi Kuantitatif Terapan*, 8(2), 136–141.
- Cortes, C., & Vapnik, V. (1995). Support-Vector Networks. *Machine Learning*, 20, 273–297.  
<https://doi.org/10.1109/64.163674>
- Hadiriyanto, I., Darsyah, M. Y., & Semarang, U. M. (2018). Peramalan Jumlah Wisatawan Mancanegara di Provinsi Bali dengan Menggunakan ARIMA dan Winter. *Prosiding Seminar Nasional Mahasiswa Unimus*, 1, 405–411.
- Hendayanti, N. P. N., Suniantara, I. K. P., & Nurhidayati, M. (2019). Penerapan Support Vector Regression (Svr) Dalam Memprediksi Jumlah Kunjungan Wisatawan Domestik Ke Bali. *Jurnal Varian*, 3(1), 43–50.  
<https://doi.org/10.30812/varian.v3i1.506>
- Herawati, S. (2016). Peramalan Kunjungan Wisatawan Mancanegara Menggunakan Generalized Regression Neural Networks. *Jurnal Infotel*, 8(1), 35–39.
- Indrasetianingsih, A., Damayanti, I., & Susanto, T. (2017). Analisis ARIMA Box Jenkins untuk Peramalan Jumlah Kunjungan Wisatawan Mancanegara di Indonesia. *Seminar Nasional Matematika Dan Aplikasinya*, 226–229.
- Lee, M. H., Rahman, N. H. A., Suhartono, Latif, M. T., Nor, M. E., & Kamisan, N. A. B. (2012). Seasonal ARIMA for Forecasting Air Pollution Index : A Case Study. *American Journal of Applied Sciences*, 9(4), 570–578.
- Makridakis, S., Wheelwright, S. C., & McGee, V. E. (1999). *Metode dan Aplikasi Peramalan (Jilid 2)*. Jakarta: Binarupa Aksara.
- Raharyani, M. P., Putri, R. R. M., & Setiawan, B. D. (2018). Implementasi Algoritme Support Vector Regression Pada Prediksi Jumlah Pengunjung Pariwisata. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 2(4), 1501–1509.
- Rosadi, D. (2012). *Ekonometrika & Analisis Runtun Waktu Terapan dengan Eviews*. Yogyakarta: Andi Offset.
- Rosadi, D. (2016). *Analisis Runtun Waktu dan Aplikasinya dengan R*. Yogyakarta: Gajah Mada University Press.
- Santoso, B. (2007). *Data Mining Teknik Pemanfaatan Data untuk Kepentingan Bisnis*. Yogyakarta: Graha Ilmu.
- Subdirektorat Statistik Pariwisata. (2017). *Neraca Satelit Pariwisata Nasional (NASPERNAS) 2017*. Retrieved from <https://www.bps.go.id/publication/2019/03/26/66604e9f077983c15b80e2bc/neraca-satelit-pariwisata-nasional--nesparnas--201>
- Wei, W. W. S. (2006). *Time Series Analysis: Univariate and Multivariate Methods Second Edition*. USA: Pearson Education.
- Widarjono, A. (2017). *Ekonometrika Pengantar dan Aplikasinya Disertai Panduan Eviews (4th ed.)*. Yogyakarta: UPP STIM YKPN.

